STRUCTURAL
BIOLOGY

Volume 74 (2018)

Supporting information for article:

MR-REX: molecular replacement by cooperative conformational search and occupancy optimization on low-accuracy protein models

Jouko J. Virtanen and Yang Zhang

Molecular replacement by cooperative conformational search and occupancy optimization on low-accuracy protein models

Jouko J. Virtanen, Yang Zhang

Supplemental Information

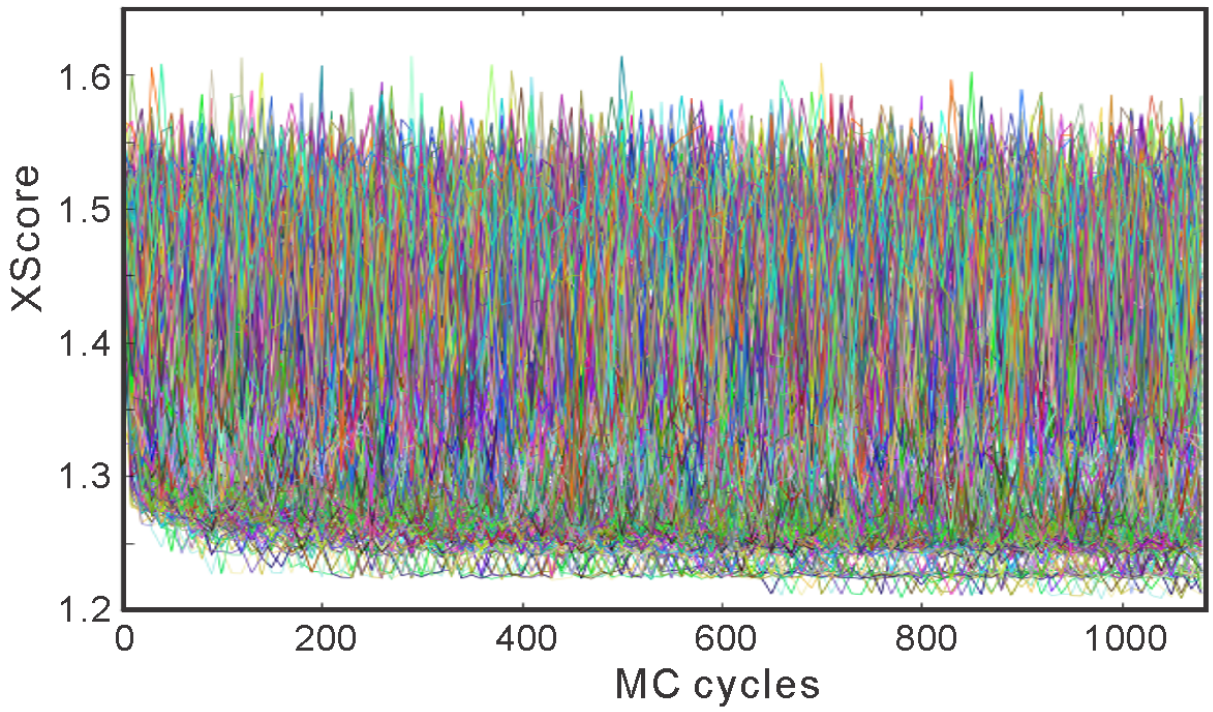
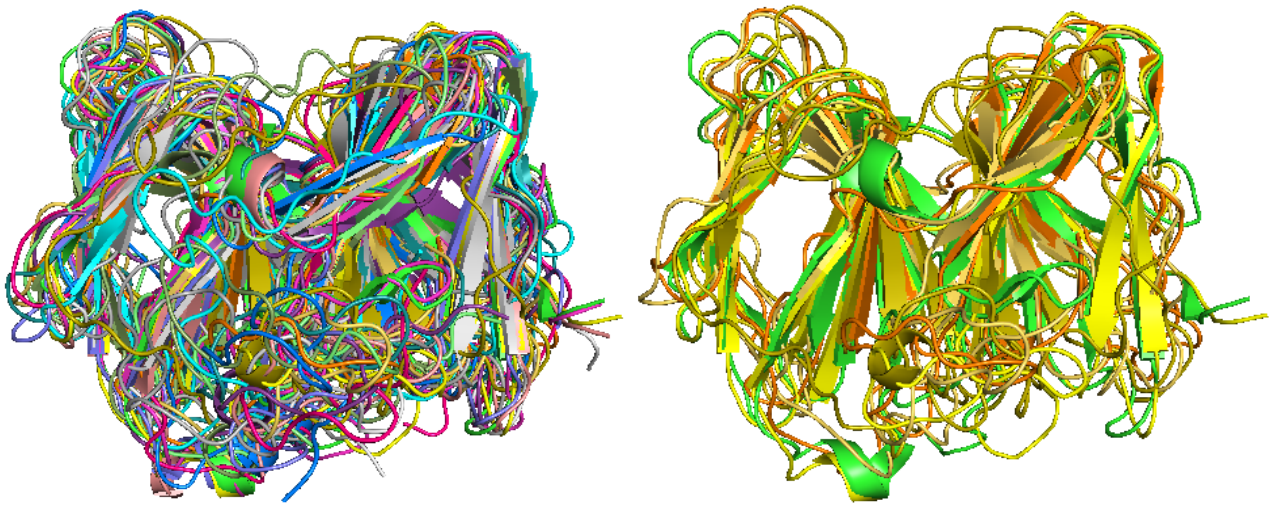
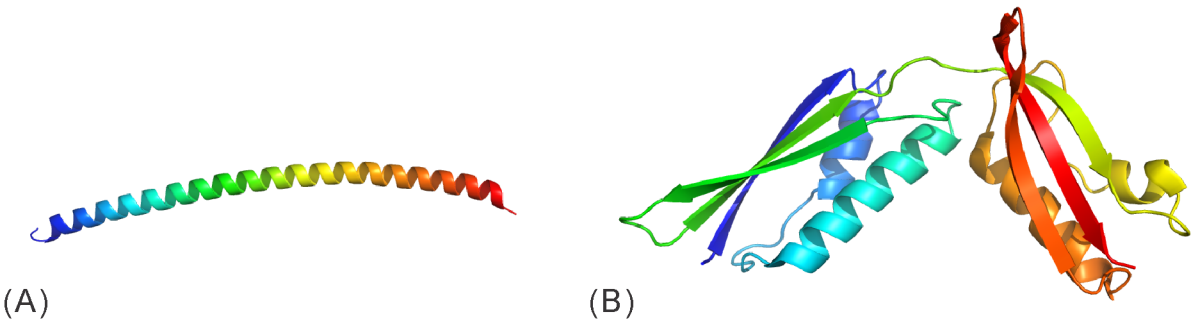


Figure S1. The *XScore* of the replicas versus the cycle in the REMC simulations. Different colors indicate the trajectories of different replicas. The trajectories are overlapped between most of the neighboring replica pairs, indicating that there are sufficient swaps between the replicas; such overlap is essential for the high-temperature replicas to help low-temperature ones to jump cross energy basins in the Replica-exchange Monte Carlo simulations. The example was taken from one of the decoys from the hexamerization domain of N-ethylmaleimide-sensitive fusion protein (PDB ID: 1D2N), where the acceptance rate of the swap movements is 98.5%.



(A) (B)

Figure S2. 2P17 is a two-domain protein. The domains of the decoys of 2P17 have the same relative orientations as the native protein. (A) All the decoys of 2P17 superimposed on the native (Green). Since it is difficult to make out the domains when all of the decoys are shown, (B) displays one fourth of decoys of 2P17 for a better view.



(A) (B)

Figure S3. Structures of the Human lamin coil 2B (A, PDB ID: 1X8Y) and *E. coli* BamA POTRA4-5 (B, PDB ID: 3Q6B). The failure point RMSD of both of these proteins is higher according to AutoBuilt R_{free} <0.4 than that according to crystallographic RMSD <2 Å, because of their approximate rotational and translational symmetries. Color blue to red runs from N- to C-terminals.

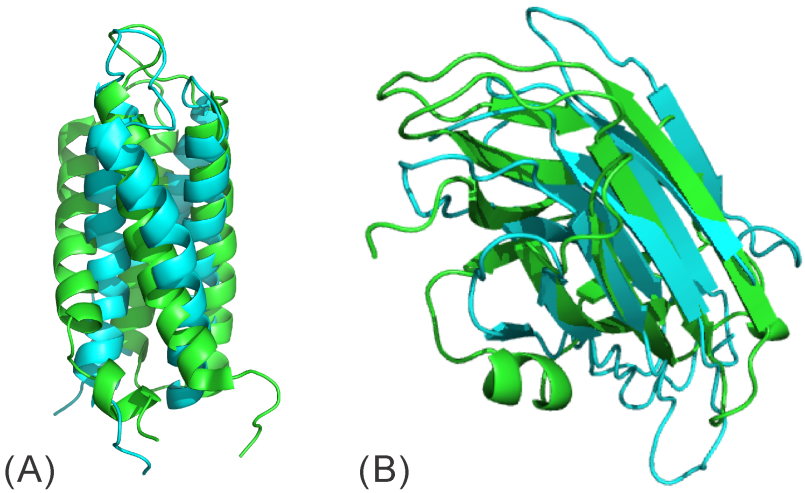


Figure S4. Two decoy structures with approximate rational symmetry that can give similar electron density as the native but are incorrect. (A) Two orientations of 2MN2 rotated 90° with respect to each other. (B) The orientation of the decoy from 2JHS (cyan) which has an electron density correlation of 0.45 to the native, despite having a 28.6 Å crystallographic RMSD, superimposed on the native (green).

Table S1. List of parameters used by MR-REX. The only parameters the user needs to specify are a, b, c, alpha, beta, gamma, SpaceGroup, XRayInput, and NumCopies.

AsaCutoff: The fraction accessible surface area below which an atom is considered to be in the core. Takes in a real number between 0 and 1. The default value is 0.2.

a: The length of unit cell vector a. Takes in a real number.

b: The length of unit cell vector b. Takes in a real number.

c: The length of unit cell vector c. Takes in a real number.

alpha: The unit cell angle alpha in degrees. Takes in a real number.

beta: The unit cell angle beta in degrees. Takes in a real number.

gamma: The unit cell angle gamma in degrees. Takes in a real number.

BulkSolventCorrection: Calculates scattering factor as Fcorrected(q)=F(q)-F(q)*ksol*exp(-bsol*q^2). Acceptable values are yes, no, true, and false. The default value is false.

CalcRotationPossibleScore: Assesses the orientation of protein by calculating the maximum and minimum scattering amplitudes, given the scattering amplitudes of the individual symmetry mates and comparing to experimental scattering amplitudes. Acceptable values are yes, no, true, and false. The default value is true.

EndIfConverged: Ends replica-exchange MR early if the search has converged. Acceptable values are yes, no, true, and false. The default value is false.

FilterOutHighResolution: Specifies that not all of the experimental data should be used, but instead Miller indexes greater than specified values should not be used. Acceptable values are yes, no, true, and false. The default value is false.

MinimizeAfterEachStep: Don't use this. It doesn't actually do anything. Acceptable values are yes, no, true, and false. The default value is false.

MinimizeEverySolution: Don't use this. It doesn't actually do anything. Acceptable values are yes, no, true, and false. The default value is false.

OptimizeBFactorScale: Optimizes a B factor scale instead of performing MR. Acceptable values are yes, no, true, and false. The default value is true.

OptimizeFormFactorScale: Optimizes a scale factor for the parameters used to calculate the scattering factors of individual atoms instead of performing MR. Acceptable values are yes, no, true, and false. The default value is true.

Phase: Calculates the structure factor of a protein, calculates the corresponding electron density, performs simple electron density modification repeatedly and compares the results to experimental data. Acceptable values are yes, no, true, and false. The default value is false.

PrintOutput: Determines if the calculated X-ray diffraction data should be outputted at the end of MR. Acceptable values are yes, no, true, and false. The default value is false.

RemoveProtein: Removes the protein from the inputted atoms, in case you want to calculate the contribution of water or ligands or something. Acceptable values are yes, no, true, and false. The default value is false.

RemoveUnknownAtoms: Removes atoms which are neither standard protein atoms nor water atoms. Acceptable values are yes, no, true, and false. The default value is false.

ReplicaExchange: Performs MR by performing a replica-exchange search. Acceptable values are yes, no, true, and false. The default value is false.

RigidBodyOptimization: Performs gradient based optimization of all replicas. Acceptable values are yes, no, true, and false. The default value is false.

RotateTranslate: Performs more traditional MR in which a rotational search is performed first followed by a translational search. Acceptable values are yes, no, true, and false. The default value is false.

SetTempBasedOnScores: Sets the temperatures of the replicas based on the initial scores of the highest and lowest scoring replicas. Acceptable values are yes, no, true, and false. The default value is false.

ShuffleDegreesOfFreedom: If there is more than one protein in the asymmetric unit cell the XRayScores for all permutations of the copies are calculated. Acceptable values are yes, no, true, and false. The default value is false.

UsyDynamicTemperature: Periodically updates the temperatures of the replicas in order to optimize the efficiency of replica-exchange. Acceptable values are yes, no, true, and false. The default value is false.

UseExpLookUp: Uses look up table for the exponential function. Acceptable values are yes, no, true, and false. The default value is false.

UseFastTranslationRotation: Uses the method of Crowther to calculate the structure factor quickly. Acceptable values are yes, no, true, and false. The default value is false.

UseFastTranslation: Calculates the effect of translating by changing the phase the structure factors of the individual symmetry mates. Acceptable values are yes, no, true, and false. The default value is false.

UseMillerIndexesFromExperiment: Calculates the structure factor for Miller indexes present in an input experimental data file, instead of all Miller indexes in a specified range. Acceptable values are yes, no, true, and false. The default value is false.

UseNCS: Enforces noncrystallographic symmetry when performing MR. Currently only works for dimers. Acceptable values are yes, no, true, and false. The default value is false.

UseVeryFastTranslation: Takes advantage of the fact that some space groups have repeated blocks of rotation matrixes that are related by translation, to reduce the number of translations that have to be calculated. Acceptable values are yes, no, true, and false. The default value is true.

Verbose: Prints more stuff than usual. Acceptable values are yes, no, true, and false. The default value is false.

ClashScoreUse: Should the clash score be used during the replica-exchange search or as a filter at the end of it. Acceptable values are DuringMR and AtEnd. The default value is AtEnd.

ContinuationFile: Outputs the degrees of freedom of the replicas to this file at the end of the replica-exchange search. Takes in a file path.

CorrectionFile: The scale factors for each q bin to get the calculated structure factor to agree with the experimental data. Takes in a file path.

CubeFile: A pdb file with the electron density of the solvent on a grid of points, given in the B factor column. Takes in a file path.

DcdFilePaths: A file containing a list of dcd file paths from which an average structure factor over a trajectory can be calculated. Takes in a file path.

DegOfFreedomFile: A file containing the degrees of freedom of the replicas from a previous replica-exchange search. This is a ContiunationFile from a previous run. Takes in a file path.

DensityOutput: File to which the calculated electron density is outputted to. Takes in a file path.

FastTranslationRotationInterpolation: The only acceptable value is Linear. Therefor this is not really used anymore.

LogFiles: This option is not used anymore.

MinimizeRotationTranslation: The type of rigid body gradient based minimization performed at the end of the replica-exchange search. Acceptable values are Analytical, RotationOnly, EstimatedGradient, ConjugateGradient, and OneAtATime. The default value is Analytical.

NativePdbFile: The correct native structure of the protein. If this file is specified the electron density correlation between the native and the model is calculated. Takes in a file path.

PdbFile: This is used when the average structure factor over a trajectory is desired and the trajectory is in dcd format and the atoms in the system need to be specified. Takes in a file path. The file needs to be in pdb format.

PdbPathsFile: A file containing a list of pdb files for which the average structure is calculated. Takes in a file path.

SpaceGroup: The space group of the crystal. It cannot have spaces.

TranslationRotationMethod: This is out of date. The acceptable values are Patterson, Phase, and RFactor. RFactor is the only one that should be used. It has a misleading name since RFactor is not neccesarily the only scoring function used when RFactor is chosen.

XRayInput: Experimental X-ray data in cif file format. Takes in a file path.

XRayOutput: The file to which the structure factor calculated by CalcXRay is outputted. Takes in a file path.

XRayScoreType: This is not really used anymore. The acceptable values are RFactor, DiffOfSquares, Pearson, Product, MLRF0, and MLTF. The default value is DiffOfSquares.

ContinuousHValues: Specifies how large the grid of Miller index vectors is when using FastTranslationRotation. Takes in an integer. The default value is 100.

ContinuousKValues: Specifies how large the grid of Miller index vectors is when using FastTranslationRotation. Takes in an integer. The default value is 100.

ContinuousLValues: Specifies how large the grid of Miller index vectors is when using FastTranslationRotation. Takes in an integer. The default value is 100.

MaxTrajectoryStructures: The maximum number of structures to use when calculating the average structure factor over a trajectory. Takes in an integer. The default value is 1000.

NumCopies: The number of copies of the protein in the asymmetric unit. Takes in an integer. The default value is 1.

NumCycles: The maximum number of replica-exchange cycles. Takes in an integer. The default is 30000.

NumQxValues: The maximum value of Miller index h. Takes in an integer. The default is 16.

NumQyValues: The maximum value of Miller index k. Takes in an integer. The default is 16.

NumQzValues: The maximum value of Miller index l. Takes in an integer. The default is 16.

NumReplicas: The number of replicas used in the replica-exchange simulation. Takes in an integer. The default is 300.

MRIterations: The number of cycles of structure factor calculations and electron density modification used. Normally does nothing. Takes in an integer. The default is 1.

RandomSeed: The random seed used in the replica-exchange search. Takes in an integer. The default value is 0.

ReplicaStrutureOutput: The prefix of the candidate MR solutions to which a suffix is added for each replica.

ReplicaXRayOutput: The prefix of the structure factor files of the candidate MR solutions to which a suffix is added for each replica.

RotationGridSearchPoints: The number of angles to search through for each axis, when doing rotational search instead of replica-exchange. Takes in an integer. The default is 31.

ShuffleInterval: This is used when there is more than one copy in the asymmetric unit cell and ShuffleDegreesOfFreedom is set to true. This determined the frequency with which the copies in each replica is permuted. Takes in an integer. The default is 1000.

StepsPerCycle: The number of Monte Carlo steps per replica per cycle of replica-exchange search. Takes in an integer. The default value is 20.

TranslationGridSearchPoints: The number of grid points in each dimension on which the protein is placed when doing a translational search. Takes in an integer. The default value is 31.

XCubes: Not used.

YCubes: Not used.

ZCubes: Not used.

BFactorScale: A scale factor which is applied to all B factors. Takes in a positive real number. The default value is 1.

BulkDensity: The electron density of the buffer in electrons per cubic angstrom. Takes in a positive real number. The default value is 0.334.

CalphaRadius: The radius of alpha carbon atoms for the purpose of calculating accessible surface area. Takes in a positive real number. The default value is 4.

ClashWeightCoreCore: The weight applied on pairs of clashing core atoms. Takes in a positive real number. The default value is 1000.

ClashWeightCoreSurface: The weight applied on core atoms clashing with surface atoms. Takes in a positive real number. The default value is 10.

ClashWeightSurfaceSurface: The weight applied on pairs of clashing surface atoms. Takes in a positive real number. The default value is 1.

CubeSize: Not used.

DiffOfSquaresWeight: The weight applied to the DiffOfSquares score. Takes in a positive real number. The default value is 0.1.

ExcludedVolumeRadii: The method by which excluded volume radii are set. The acceptable values are Default and Params. Default sets the excluded volume radii to default values while Params sets them to values specified in the parameter file. The default value is Default.

ExcludedVolumeRadiiusHydrogen: The excluded volume of hydrogen atoms. Takes in a positive real number. The default value is 0.90225.

ExcludedVolumeRadiiusCarbon: The excluded volume of carbon atoms. Takes in a positive real number. The default value is 1.43693.

ExcludedVolumeRadiiusNitrogen: The excluded volume of nitrogen atoms. Takes in a positive real number. The default value is 1.24527.

ExcludedVolumeRadiiusOxygen: The excluded volume of oxygen atoms. Takes in a positive real number. The default value is 1.22099.

ExcludedVolumeRadiiusSulfur: The excluded volume of sulfur atoms. Takes in a positive real number. The default value is 2.19596.

scale: A scale factor applied to the exponentials used in calculating the structure factors of the individual atoms. This should always be set to 1. This is only used for testing. Takes in a positive real number. The default value is 1.

LogSquareDiffWeight: The weight applied to the LogSquareDiff score. Takes in a positive real number. The default value is 0.

MaxAngleChange: The initial width, in radians, of the Gaussian distribution which determines the probability of making a rotation of a certain size of the highest temperature replica. Takes in a positive real number. The default value is 0.12.

MaxReplicaExchangeTime: The maximum time, in hours, that the replica-exchange search lasts. Takes in a positive real number. The default value is 4.

MaxTemp: The temperature of the highest temperature replica, when the temperatures are not set according to the initial XRayScores. Takes in a positive real number. The default value is 0.3.

MaxTempFrac: The fraction of the highest initial XRayScore to which the temperature of the highest temperature replica is set.

MaxTranslationFraction: The initial width, in fractional coordinates, of the Gaussian distribution which determines the probability of making a translation of a certain size of the highest temperature replica. Takes in a positive real number. The default value is 0.2.

MaxQx: Not used.

MaxQy: Not used.

MaxQz: Not used.

MinAngleChange: The initial width, in radians, of the Gaussian distribution which determines the probability of making a rotation of a certain size of the lowest temperature replica. Takes in a positive real number. The default value is 0.05.

MinTemp: The temperature of the lowest temperature replica, when the temperatures are not set according to the initial XRayScores. Takes in a positive real number. The default value is 0.02.

MinTempFrac: The fraction of the lowest initial XRayScore to which the temperature of the lowest temperature replica is set. Takes in a positive real number. The default value is 0.005.

MinTranslationFraction: The initial width, in fractional coordinates, of the Gaussian distribution which determines the probability of making a translation of a certain size of the lowest temperature replica. Takes in a positive real number. The default value is 0.1.

MLRF0Weight: The weight applied to the maximum likelihood rotation function. Takes in a positive real number. The default value is 0.

PearsonWeight: The weight applied to the pearson scoreing function. Takes in a positive real number. The default value is 1.

ProductWeight: The weight applied to the sum of the products of the calculated and observed reflection intensities. Takes in a negative real number. The default value is 0.

RFactorWeight: The weight applied to the R factor. Takes in a positive real. The default value is 1.

RotationWeight: The weight applied to the rotation possible score. Takes in a positive real number. The default value is 0.

XCubeLength: Not used.

YCubeLength: Not used.

ZCubeLength: Not used.

Table S2. The data used in the linear regressions for MR-REX.

Protein (PDB ID)	Volume[a]	Density[b]	ε[c]	n_{sym}[d]	n_{het}[e]	$n_{reflection}$[f]	TM-score Phaser[g]	TM-score MR-REX[h]
1D2N	471598	0.00052	0.282	6	0.158	5982	0.83	0.86
1OKC	355985	0.00082	0.145	4	1.195	3649	0.76	0.81
1R0U	276542	0.00053	0.150	6	0.049	3408	0.84	0.84
1SU0	584531	0.00023	0.132	16	0.014	1611	0.79	0.76
1V05	235388	0.01159	0.280	12	0.010	844	0.77	0.77
1VPQ	584113	0.00044	0.122	8	0.061	2860	0.82	0.80
1X8Y	521640	0.00014	0.739	12	0.014	3177	0.81	0.77
2BOU	198833	0.00071	0.548	4	0.073	2489	0.76	0.87
2IL5	392004	0.00042	0.196	6	0.012	4503	0.74	0.64
2P17	223325	0.00111	0.146	4	0.008	2292	0.77	0.88
2RBK	262472	0.00099	0.188	4	0.022	5028	0.79	0.90
2YQ9	208721	0.00104	0.208	4	0.168	2489	0.74	0.77
3B7C	478329	0.00025	0.153	12	0.193	3210	0.83	0.75
3BW6	348506	0.00039	0.144	6	0.044	2022	0.72	0.66
3CHV	250134	0.00111	0.060	4	0.017	2490	0.74	0.61
3FZX	884993	0.00024	0.102	12	0.014	5092	0.83	0.83
3HYN	162705	0.00114	0.125	4	0.042	1851	0.86	0.79
3K93	470647	0.00047	0.230	6	0.105	5471	0.74	0.77
3MT0	130007	0.00223	0.240	2	0.014	2752	0.81	0.81
3N2Q	523874	0.000546	0.313	6	0.004	6132	0.85	0.86
3ONJ	82580.2	0.00117	0.390	4	0.010	1033	0.80	0.77
3PU6	152889	0.00096	0.128	4	0.007	1958	0.76	0.69
3PYW	607238	0.00029	0.135	8	0.034	3333	0.70	0.72
3Q6B	171460	0.00090	0.375	4	0.006	1829	0.62	0.62
3VQF	89313.6	0.00095	0.153	4	0.011	1184	0.70	0.70
3VWC	163822	0.00089	0.125	4	0.059	1702	0.76	0.79
3ZDB	348264	0.00071	0.240	4	0.036	3772	0.79	0.76
4A3Z	312765	0.00043	0.171	8	0.014	1838	0.80	0.79
4DCD	354951	0.00052	0.093	8	0.237	1836	0.78	0.71
4IS7	960267	0.00015	0.413	12	0.007	5282	0.80	0.78
4KR1	872071	0.00028	0.247	12	0.005	5116	0.80	0.80
4L8G	230910	0.00072	0.106	6	0.265	1353	0.75	0.75
4LVP	471336	0.00027	0.158	12	0.040	3019	0.77	0.70
4M6T	1.80843e6	0.00010	0.220	18	0.158	6614	0.70	0.74
4MDN	338157	0.00028	0.162	8	0.526	1758	0.64	0.65
4MJF	304617	0.00075	0.251	4	0.317	3051	0.79	0.75
4NBR	481647	0.00056	0.132	8	0.024	2558	0.88	0.84
4OQ4	361529	0.00052	0.054	6	0.052	3925	0.80	0.79

[a]The volume of the unit cell.
[b]The packing density of the unit cell in terms of residues per $Å^3$.
[c]The elongation rate of protein structure.
[d]The number of symmetry mates in the unit cell.
[e]The number of HETATM per residue.
[f]The number of reflections used for MR.
[g]The TM-score of the worst decoy structure succeed by Phaser.
[h]The TM-score of the worst decoy structure succeed by MR-REX.